



**HAL**  
open science

# Robust risk management via multi-marginal optimal transport

Quentin Mérigot, Luca Nenna, Hamza Ennaji, Brendan Pass

► **To cite this version:**

Quentin Mérigot, Luca Nenna, Hamza Ennaji, Brendan Pass. Robust risk management via multi-marginal optimal transport. 2022. hal-03852245

**HAL Id: hal-03852245**

**<https://hal.science/hal-03852245>**

Preprint submitted on 14 Nov 2022

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Robust risk management via multi-marginal optimal transport <sup>\*</sup>

Hamza Ennaji<sup>†</sup>    Quentin Mérigot<sup>‡</sup>    Luca Nenna<sup>§</sup>  
Brendan Pass<sup>¶</sup>

November 14, 2022

## Abstract

We study the problem of maximizing a spectral risk measure of a given output function which depends on several underlying variables, whose individual distributions are known but whose joint distribution is not. We establish and exploit an equivalence between this problem and a multi-marginal optimal transport problem. We use this reformulation to establish explicit, closed form solutions when the underlying variables are one dimensional, for a large class of output functions. For higher dimensional underlying variables, we identify conditions on the output function and marginal distributions under which solutions concentrate on graphs over the first variable and are unique,

---

<sup>\*</sup>The work of BP was partially supported by National Sciences and Engineering Research Council of Canada Discovery Grant number 04658-2018. He is also pleased to acknowledge the generous hospitality of the Institut de Mathématiques d’Orsay, Université Paris-Sud during his stay in November of 2019 as a missionnaire scientifique invité, when this work was partially completed. L.N. was supported by a public grant as part of the Investissement d’avenir project, reference ANR-11-LABX-0056-LMH, LabEx LMH, from the CNRS PEPS JCJC (2022) and H-CODE.

<sup>†</sup>Normandie Université, ENSICAEN, UNICAEN, CNRS, GREYC, France. [hamza.ennaji@unicaen.fr](mailto:hamza.ennaji@unicaen.fr)

<sup>‡</sup>Université Paris-Saclay, CNRS, Laboratoire de mathématiques d’Orsay, 91405, Orsay, France. [quentin.merigot@universite-paris-saclay.fr](mailto:quentin.merigot@universite-paris-saclay.fr)

<sup>§</sup>Université Paris-Saclay, CNRS, Laboratoire de mathématiques d’Orsay, 91405, Orsay, France. [luca.nenna@universite-paris-saclay.fr](mailto:luca.nenna@universite-paris-saclay.fr)

<sup>¶</sup>Department of Mathematical and Statistical Sciences, 632 CAB, University of Alberta, Edmonton, Alberta, Canada, T6G 2G1 [pass@ualberta.ca](mailto:pass@ualberta.ca)

and, for general output functions, we find upper bounds on the dimension of the support of the solution. We also establish a stability result on the maximal value and maximizing joint distributions when the output function, marginal distributions and spectral function are perturbed; in addition, when the variables are one dimensional, we show that the optimal value exhibits Lipschitz dependence on the marginal distributions for a certain class of output functions. Finally, we show that the equivalence to a multi-marginal optimal transport problem extends to maximal correlation measures of multi-dimensional risks; in this setting, we again establish conditions under which the solution concentrates on a graph over the first marginal.

## Contents

<b>1</b>	<b>Introduction</b>	<b>2</b>
<b>2</b>	<b>Preliminaries and problem formulation</b>	<b>6</b>
2.1	Maximal spectral risk measures over couplings of given marginals	11
<b>3</b>	<b>Equivalence between maximizing spectral risk measures and multi-marginal optimal transport</b>	<b>14</b>
<b>4</b>	<b>Solutions for one-dimensional assets and compatible payouts</b>	<b>17</b>
4.1	Sensitivity analysis and maximal river flow . . . . .	21
<b>5</b>	<b>Higher dimensional assets</b>	<b>22</b>
<b>6</b>	<b>Stability</b>	<b>25</b>
<b>7</b>	<b>Multidimensional measures of risk</b>	<b>28</b>

## 1 Introduction

In a variety of problems in operations research, a variable of interest  $b = b(x_1, x_2, \dots, x_d)$  depends on several underlying random variables, whose individual distributions are known (or can be estimated) but whose joint distribution is not. A natural example arises in finance, when one considers the payout of a derivative depending on several underlying assets. An estimate

of the distribution of the asset values themselves can often be inferred from the prices of vanilla call and put options for a wide range of strike prices; since these options are widely traded, their prices are readily available. However, estimating the joint distribution would require prices of a wide range of derivatives with payouts depending on all the variables, which are typically much scarcer. Other examples arise when the variables represent parameters in a physical system whose individual distributions can be estimated empirically or through modeling (or a combination of both) but whose dependence structure cannot; an example of this flavour, originating in [18], in which the output  $b$  represents the simulated height of a river at risk of flooding, and the underlying variables include various design parameters and climate dependent factors can be found in Section 4.1 below. A third example comes from election projections in political science, where  $b$  might represent the outcome of an election and the  $x_i$  vote shares in different regions, whose individual distributions can be modelled from polling data.

Methods used in risk management to evaluate the resulting aggregate level of risk naturally depend on the distribution of the output variable  $b$ , and therefore, in turn, on the joint distribution of the  $x_i$ . A natural problem is therefore to determine bounds, or worst-case scenarios, on these risk measures; that is, to maximize a given risk measure over all possible joint distributions of the  $x_i$  with known marginal distributions. Problems such as this have received extensive attention within the financial risk management community; see, for example, [32]. In the simplest of these problems, the underlying variables  $x_i$  are typically real valued and the output variable  $b$  is often assumed to have a particular structure (in many cases, it is a weighted sum of the  $x_i$ , reflecting the value of a portfolio built out of underlying assets with values  $x_i$ , or a function of this weighted sum). In these cases, explicit solutions for the maximizing couplings of the  $x_i$  can sometimes be obtained (see, for example, [29, 30, 36]).

For more general output functions  $b$ , and possibly multi-dimensional underlying variables  $x_i$ , much less is known about the dependence structure of the maximizing joint distributions. A recent paper of Ghossoub-Hall-Saunders [15] studies this more general setting systematically, allowing the underlying variables to take values in very general spaces and the output  $b$  to take a very general form, and focuses on *spectral risk measures* (see Definition 10 below). They observe that the resulting problem is a generalization of optimal transport (see Section 2 for a brief overview of the optimal transport problem); in fact, in the simplest case, when the spectral function is identi-

cally equal to one and the number  $d$  of underlying variables is 2, the problem is exactly a classical optimal transport problem with surplus function  $b$ . For more general spectral risk measures, the problem has the *same constraint set* as the optimal transport problem, but a more general, *non-linear objective functional*; they adapt the duality theory of optimal transport to this setting and establish results on the stability with respect to perturbations of  $b$  and the marginal distributions. Although the analysis in [15] focused on the  $d = 2$  case, they note that their results can be extended to the  $d \geq 3$  setting, in which case maximizing spectral risks becomes a generalization of the multi-marginal optimal transport problem (see Section 2 for a brief overview of this problem). As in the two marginal case, one maximizes a concave objective function over the set of couplings of the given marginals – when the spectral function is constant, one obtains multi-marginal optimal transport, a linear maximization.

The main purpose of this paper is to establish and exploit a simpler, but equivalent, reformulation of this problem. Specifically, we show that for *any* spectral risk measure, the maximization can in fact be formulated as a traditional multi-marginal optimal transport problem with  $d + 1$  marginals: the given marginals distributions of the  $x_i$  as well as another distribution arising from the particular form of the spectral function. Although this formulation slightly increases the underlying dimension, it makes the maximization problem linear and much more tractable – indeed, the results and techniques, both theoretical and computational, in the substantial literature on multi-marginal optimal transport become directly applicable.

Though the structure of solutions to multi-marginal optimal transport problems is in general a notoriously delicate issue (see [26] for an overview) there are many important cases when the problem is well understood. As a consequence of our result, when the underlying variables  $x_i$  are all one dimensional, we derive an explicit characterization of solutions for a substantial class of output functions  $b$ , through a careful refinement of the existing theory of multi-marginal optimal transport on one dimensional ambient spaces. For  $b$  falling outside this class, explicit solutions are likely generally unattainable; however, our formulation of the problem can potentially facilitate the use of a very broad range of computational methods for optimal transport to approximate solutions numerically (see Section 8.1 in [2] and the references therein).

For underlying variables in higher dimensional spaces, explicit solutions are generally not possible. However, we show that known techniques from

multi-marginal optimal transport can be adapted to identify conditions under which the solution concentrates on a graph over the first variable; solutions having this structure, commonly referred to as *Monge solutions* in the optimal transport literature, are analogous to the well known comonotone couplings which often maximize spectral risk measures in the one dimensional case, since they essentially assert that knowing the value of the first variable completely determines the values of the others. On the other hand, for many output functions  $b$ , conditions ensuring this structure of a solution do not hold; in this case, we demonstrate that available techniques can still be used bound the dimension of the support of the optimizer. In addition to being of theoretical interest, this fact is potentially important in future work on the computation of solutions.<sup>1</sup>

We go on to show that our formulation, combined with standard stability results for optimal transport problems, implies that the maximal value of the spectral risk measure, and the maximizing joint distributions, are stable with respect to perturbations of the marginals, output function and spectral function. When the underlying variables are one dimensional and the output function satisfies additional hypotheses, we further leverage the connection to optimal transport to show that the maximal value's dependence on the marginals is in fact Lipschitz.

We take particular note of the special but important case when the spectral risk measure in question is the *expected shortfall*; in this case, the problem further reduces to a *multi-marginal partial transport problem* on the  $d$  original distributions. In this case, in higher dimensions, results in [6, 12] and [20] immediately apply, implying uniqueness and graphical structure of solutions for a certain class of output functions  $b$ . Furthermore, the fact, exposed by our work here, that optimal partial transport can be reformulated as a multi-marginal optimal transport problem does not seem to have been observed before, and may be of independent interest in the optimal transport

---

<sup>1</sup>In general, the joint distributions of interest are probability measures on an  $n \cdot d$  dimensional space, where  $d$  is the number of variables, and  $n$  the dimension of each variable (assuming, for simplicity of exposition, that all variables are of the same dimension). When maximizers concentrate on graphs over  $x_1$ , they are essentially much simpler,  $n$ -dimensional objects. More generally, it is possible to prove that optimizers concentrate on  $m$  dimensional subsets of the  $n \cdot d$  dimensional ambient space, where  $m$  satisfies  $n \leq m \leq n \cdot d$  and is determined by the output function  $b$ ; this has consequences for the covering number of the support, which will potentially play an important role in future computation.

community.

We also extend these ideas to the setting where the output function  $b$  is multi-variate valued, using the maximal correlation risk measures developed in [11]. We are again able to identify conditions under which the solution concentrates on a graph.

The paper is organized as follows: in Section 2, we first introduce the traditional (two marginal) optimal transport problem and its multi-marginal generalization, and establish certain preliminary results which we will need. We then introduce spectral risk measures, and the problem of maximizing them over joint distributions with fixed marginals, which is the main object of interest in this paper (see (13) below). In Section 3, we establish the equivalence between this maximization problem and multi-marginal optimal transport. In Section 4, we establish explicit solutions for the maximizing joint distributions when the underlying marginal variables are one dimensional and in the fifth section, we study how these result generalize to the setting where the underlying variables are higher dimensional. Section 6 is focused on the stability of the problem with respect to variations in the marginals, output function and spectral risk measure, while the final section is reserved for the extension of some of our results to maximal correlation risk measures on multi-variate output functions.

## 2 Preliminaries and problem formulation

**Assumption and notations** Let  $\mu_i \in \mathcal{P}(X_i)$  be probability measures on Polish spaces  $X_i$ , for  $i = 0, 1, \dots, d$ . Given a mapping  $T : X_i \rightarrow X_j$ , we say that  $T$  pushes  $\mu_i$  forward to  $\mu_j$ , and write

$$T_{\#}\mu_i = \mu_j, \text{ if } \mu_i(T^{-1}(A)) = \mu_j(A) \text{ for all Borel } A \subseteq X_j.$$

We will let  $\Gamma(\mu_0, \mu_1, \dots, \mu_d) \subset \mathcal{P}(X_0 \times X_1 \times \dots \times X_d)$  denote the set of probability measures on  $X_0 \times X_1 \times \dots \times X_d$  whose marginals are the  $\mu_i$ ; that is,  $\mu_i = \left( (x_0, x_1, x_2, \dots, x_d) \mapsto x_i \right)_{\#} \pi$  for each  $i$  and each  $\pi \in \Gamma(\mu_0, \mu_1, \dots, \mu_d)$ .

**Definition 1** ( $(1, \dots, d)$ -marginal). *For  $\pi \in \mathcal{P}(X_0 \times X_1 \times \dots \times X_d)$ , we call the  $(1, \dots, d)$ -marginal of  $\pi$ , the projection of  $\pi$  onto  $X_1 \times \dots \times X_d$ ; that is,  $\left( (x_0, x_1, x_2, \dots, x_d) \mapsto (x_1, x_2, \dots, x_d) \right)_{\#} \pi$ . Note that if  $\pi \in \Gamma(\mu_0, \mu_1, \dots, \mu_d)$  then clearly its  $(1, \dots, d)$ -marginal is in  $\Gamma(\mu_1, \dots, \mu_d)$ .*

As usual we define the cumulative distribution function (c.d.f) of a probability measure  $\mu \in \mathcal{P}(\mathbb{R})$  as  $F_\mu(x) = \mu((-\infty, x])$ . The generalized or pseudo-inverse of  $F_\mu$  is defined as

$$F_\mu^{-1}(m) = \inf\{x \in \mathbb{R} \mid F_\mu(x) \geq m\}.$$

We recall that  $F_\mu^{-1}(m)$  is the value of the  $m$ th quantile.

**Optimal Transport problem** Given two probability distributions  $\mu_0$  and  $\mu_1$  on Polish spaces  $X_0$  and  $X_1$ , respectively, and a surplus function  $s : X_0 \times X_1 \rightarrow \mathbb{R}$ , the optimal transport problem (in its Monge formulation) consists in maximising

$$\int_{X_0} s(x, T(x)) d\mu_0(x) \tag{1}$$

under the constraint that  $T_\# \mu_0 = \mu_1$  (namely  $\mu_1$  is the image measure of  $\mu_0$  through the map  $T$ ). This is a delicate problem since the mass conservation constraint  $T_\# \mu_0 = \mu_1$  is highly nonlinear (and the feasible set may even be empty, for instance if  $\mu_0$  is a Dirac mass and  $\mu_1$  is not). For this reason, it is common to study a relaxed formulation of (1) which allows mass splitting; that is,

$$\max_{\pi \in \Gamma(\mu_0, \mu_1)} \int_{X_0 \times X_1} s(x, y) d\pi(x, y) \tag{2}$$

where, as above,  $\Gamma(\mu_0, \mu_1)$  consists of all probability measures on  $X_0 \times X_1$  having  $\mu_0$  and  $\mu_1$  as marginals. Note that this is a (infinite dimensional) linear programming problem and that there exists solutions under very mild assumptions on  $s$ ,  $\mu_0$  and  $\mu_1$ . A solution  $\pi$  of (2) is called an optimal transport plan; in particular, if an optimal plan of (2) has a deterministic form  $\pi = (\text{id}, T)_\# \mu_0$  (which means that no splitting of mass occurs and  $\pi$  is concentrated on the graph of  $T$ ) then  $T$  is an optimal transport map, i.e., a solution to (1). It is therefore sometimes called a *Monge solution* of (2). The linear problem (2) also has a convenient dual formulation

$$\min_{\substack{u, v \\ u(x) + v(y) \geq s(x, y)}} \int_{X_0} u(x) d\mu_0(x) + \int_{X_1} v(y) d\mu_1(y) \tag{3}$$

where  $u(x)$  and  $v(y)$  are the so-called Kantorovich potentials. Optimal Transport theory for two marginals has developed very rapidly in the 25 last years; there are well known conditions on  $s$ ,  $\mu_0$  and  $\mu_1$  which guarantee that there is a unique deterministic optimal plan and we refer to the textbooks by Santambrogio [33] and Villani [34, 35], for a detailed exposition.



**Multi-marginal optimal transport** In this paper we consider a generalization of (1) to the case in which more than two marginals are involved: given probability measures  $\mu_0, \mu_1, \dots, \mu_d$  on spaces  $X_0, X_1, \dots, X_d$ , and a surplus function  $s(x_0, x_1, \dots, x_d)$ , the multi-marginal optimal transport problem is to maximize

$$\int_{X_0 \times X_1 \times \dots \times X_d} s(x_0, x_1, \dots, x_d) d\pi(x_0, x_1, \dots, x_d) \quad (4)$$

over  $\pi \in \Gamma(\mu_0, \mu_1, \dots, \mu_d)$ . This problem has attracted increasing attention in recent years since it arises naturally in many different settings such as economics [10], quantum chemistry [5, 9], fluid mechanics [4], etc. Our results in Section 3 will bring forth a new application to maximizing spectral risk measures.

It is well known that under mild conditions (for example, continuity of  $s$  and compact support of the  $\mu_i$  is more than sufficient) (4) admits a solution; see, for example, [26]. Although the structure of solutions to (4) can generally depend delicately on  $s$ , there is a growing theory, and an important class for which solutions can be derived essentially explicitly, which we describe below. In addition, there are now powerful numerical tools to compute solutions for problems falling outside this class (for a more detailed discussion, we refer to Section 8.1 in [2] and the references therein).

The most tractable of these problems occurs when the underlying asset spaces are one dimensional, and the mixed second derivatives of  $s$  interact in a certain way.

**Definition 2** (Compatibility). *Suppose each  $X_i \subset \mathbb{R}$  is a bounded real interval. Assume that  $s \in \mathcal{C}^2(X_0 \times X_1 \times \dots \times X_d)$ . We say that  $s$  is strictly compatible if for each three distinct indices  $i, j, k \in \{0, 1, \dots, d\}$  and each  $(x_0, \dots, x_d) \in X_0 \times \dots \times X_d$  we have*

$$\frac{\partial^2 s}{\partial x_i \partial x_j} \left[ \frac{\partial^2 s}{\partial x_k \partial x_j} \right] \frac{\partial^2 s}{\partial x_k \partial x_i}(x_0, \dots, x_d) > 0. \quad (5)$$

*We will say that  $s$  is weakly compatible (or simply compatible) if for each  $i \neq j$ , we have either  $\frac{\partial^2 s}{\partial x_i \partial x_j} \geq 0$  throughout  $X_0 \times X_1 \times \dots \times X_d$  or  $\frac{\partial^2 s}{\partial x_i \partial x_j} \leq 0$  throughout  $X_0 \times X_1 \times \dots \times X_d$ , and*

$$\frac{\partial^2 s}{\partial x_i \partial x_j} \left[ \frac{\partial^2 s}{\partial x_k \partial x_j} \right] \frac{\partial^2 s}{\partial x_k \partial x_i}(x_0, \dots, x_d) \geq 0, \quad (6)$$

for each distinct  $i, j, k \in \{0, 1, \dots, d\}$  and each  $(x_0, \dots, x_d) \in X_0 \times \dots \times X_d$ .

The fact that the mixed partials  $\frac{\partial^2 s}{\partial x_i \partial x_j}$  do not change signs under the compatibility condition allows us to partition the set  $\{0, 1, 2, \dots, d\} = S_+ \cup S_-$  of indices into disjoint subsets  $S_+$  and  $S_-$  such that  $0 \in S_+$  and for each  $i \neq j$ ,  $\frac{\partial^2 s}{\partial x_i \partial x_j} \geq 0$  throughout  $X_0 \times X_1 \times \dots \times X_d$  if either both  $i$  and  $j$  are in  $S_+$  or if both are in  $S_-$ , and  $\frac{\partial^2 s}{\partial x_i \partial x_j} \leq 0$  throughout  $X_0 \times X_1 \times \dots \times X_d$  otherwise. The inequalities are strict if strict compatibility holds. A well known special case of compatibility is captured by the following definition.

**Definition 3** (Supermodularity). *Suppose each  $X_i \subset \mathbb{R}$  is a connected real interval. Assume that  $s \in \mathcal{C}^2(X_0 \times X_1 \times \dots \times X_d)$ . We say that  $s$  is supermodular if  $\frac{\partial^2 s}{\partial x_i \partial x_j} \geq 0$  for all  $i \neq j$ ; we say  $s$  is strictly supermodular if  $\frac{\partial^2 s}{\partial x_i \partial x_j} > 0$  for all  $i \neq j$*

**Remark 4.** *Notice that if  $s$  is supermodular, then it is clearly compatible. On the other hand, if  $s$  is compatible, define  $\tilde{s}(\tilde{x}_0, \tilde{x}_1, \dots, \tilde{x}_d) = s(x_0, x_1, \dots, x_d)$ , where  $\tilde{x}_0 = x_0$  and for each  $i = 1, 2, \dots, d$  we set  $\tilde{x}_i = x_i$  if  $i \in S_+$  and  $\tilde{x}_i = -x_i$  if  $i \in S_-$ . Then it is straightforward to show that  $\tilde{s}$  is supermodular. Therefore, compatibility is simply supermodularity up to a change of variables.*

*Both supermodularity and compatibility may be extended to non  $C^2$  functions, by considering partial differences; for simplicity of exposition, we stick with the version relying on second derivatives here.*

We then introduce the following special coupling in  $\Gamma(\mu_0, \mu_1, \dots, \mu_d)$ .

**Definition 5.** *For a compatible  $s$ , we define the  $s$ -comonotone coupling by:*

$$\pi = (G_0, G_1, \dots, G_d)_{\#} \text{Leb}_{[0,1]}, \quad (7)$$

where  $G_0 = F_{\mu_0}^{-1}$  and for each  $i = 1, 2, \dots, d$

$$G_i(m) = \begin{cases} F_{\mu_i}^{-1}(m) & \text{if } i \in S_+, \\ F_{\mu_i}^{-1}(1 - m) & \text{if } i \in S_-. \end{cases} \quad (8)$$

2

---

<sup>2</sup>There may be multiple ways to define the  $G_i$  (if, for example, for some  $i$ ,  $\frac{\partial^2 s}{\partial x_i \partial x_j} = 0$  everywhere for all  $j$ ); in this case, the results below apply to all resulting such  $s$ -comonotone couplings. Note, however, that the  $G_i$  are uniquely defined when compatibility is strict.

**Remark 6.** *If  $s$  is supermodular, we can take  $S_-$  to be empty and  $S_+ = \{0, 1, \dots, n\}$ ; the  $s$ -comonotone coupling is then exactly the well known classical comonotone coupling.*

The following result can be found in [7] for strictly supermodular functions as well as in [22] for compatible functions (where its proof appears together with the formulation of the compatibility conditions and the observation that it is equivalent to strict supermodularity after changing coordinates).

**Theorem 7.** *Suppose that  $s$  is strictly compatible. Then the  $s$ -comonotone coupling (7) is the unique optimizer in (4).*

**Remark 8.** *A classical but important case occurs when  $d = 2$  and  $s(x_0, x_1) = x_0x_1$ . It is well known that there is only one measure in  $\Gamma(\mu_0, \mu_1)$  with monotone increasing support. This will be used in the proof of Lemma 16 below.*

*For  $d \geq 3$ , the  $s$ -comonotone  $\pi$  is characterized by the fact that its twofold marginals,  $\pi_{ij} = ((x_0, x_1, \dots, x_d) \mapsto (x_i, x_j))_{\#}\pi$  are the monotone increasing (respectively decreasing) couplings if  $\frac{\partial^2 s}{\partial x_i \partial x_j} > 0$  (respectively  $\frac{\partial^2 s}{\partial x_i \partial x_j} < 0$ ); compatibility ensures existence of a  $\pi$  with this property.*

Though the following result does not seem to be available in the literature, it is a straightforward consequence of Theorem 7 and the well known stability of optimal transport with respect to perturbations of the cost function. It asserts that the optimality of the measure constructed in Theorem 7 still holds if the strong compatibility assumption is relaxed to weak compatibility, although the uniqueness assertion may not.

**Proposition 9.** *Suppose that  $s$  is weakly compatible. Then the  $s$ -comonotone coupling (7) is optimal in (4).*

*Proof.* For  $\epsilon > 0$ , define

$$s_\epsilon(x_0, x_1, \dots, x_d) = s(x_0, x_1, \dots, x_d) + \epsilon \sum_{i,j \in S_+} (x_i - x_j)^2 + \epsilon \sum_{i,j \in S_-} (x_i - x_j)^2 - \epsilon \sum_{i \in S_+, j \in S_-} (x_i - x_j)^2$$

notice now that  $s_\epsilon$  is strictly compatible and so Theorem 7 implies that  $\pi = (G_0, G_1, \dots, G_d)_{\#}\text{Leb}_{[0,1]}$  is optimal in (4) for each  $\epsilon$ . Therefore, for any other  $\tilde{\pi} \in \Gamma(\mu_0, \mu_1, \dots, \mu_d)$  we have

$$\int s_\epsilon d\tilde{\pi} \leq \int s_\epsilon d\pi$$

Since the  $s_\epsilon$  converge uniformly to  $s$  as  $\epsilon \rightarrow 0$ , we can take the limit in the inequality above and obtain

$$\int s d\tilde{\pi} \leq \int s d\pi.$$

As  $\tilde{\pi} \in \Gamma(\mu_0, \mu_1, \dots, \mu_d)$  was arbitrary, this yields the desired result.  $\square$

We also note that, as for the two marginal problem, (4) admits a very useful dual formulation:

$$\inf \left\{ \sum_{i=0}^d \int_{X_i} u_i(x_i) d\mu_i(x_i) \mid u_i \in \mathcal{C}_b(X_i), \forall i = 0, \dots, d, \right. \\ \left. \sum_{i=0}^d u_i(x_i) \geq s(x_0, \dots, x_d) \right\}. \quad (9)$$

As with the primal problem (4), it is well known that under mild conditions a solution  $(u_0, u_1, \dots, u_d)$  to (9) exists; see, for example, [14, 8, 23]. In particular, given optimal solutions  $\pi$  and  $(u_0, \dots, u_d)$  to (4) and (9), respectively, the following optimality condition holds

$$\sum_{i=0}^d u_i(x_i) = s(x_0, \dots, x_d), \quad \pi - \text{a.e.} \quad (10)$$

moreover, by Proposition 9 for a compatible  $s$  (10) rewrites

$$\sum_{i=0}^d u_i(G_i(m)) = s(G_0(m), \dots, G_d(m)), \quad \text{Leb} - \text{a.e.}$$

## 2.1 Maximal spectral risk measures over couplings of given marginals

Let the real line describe a certain level of risk (e.g., the level of radiation in the nuclear power plant) and  $\mu$  be a probability measure on  $\mathbb{R}$  which can be interpreted as the distribution of risk. We will consider the following form of quantifier of the risk associated with  $\mu$ .

**Definition 10** (spectral risk measure). A functional  $\mathcal{R} : \mathcal{P}(\mathbb{R}) \rightarrow \mathbb{R}$  is called a spectral risk measure if it takes the form  $\mathcal{R} = \mathcal{R}_\alpha$ , where

$$\mathcal{R}_\alpha(\mu) := \int_0^1 F_\mu^{-1}(m)\alpha(m)dm. \quad (11)$$

**Remark 11.** Note that by taking  $\alpha$  to be defined and real valued on  $[0, 1]$ , we are tacitly assuming that it is bounded, since  $\alpha(0) \leq \alpha(m) \leq \alpha(1)$  by monotonicity; the same assumption was made in [15]. This is mostly for technical convenience; we expect that most of our results can be extended to the more general case  $\alpha : [0, 1] \rightarrow \mathbb{R}_+$ , where possibly  $\lim_{m \rightarrow 1} \alpha(m) = \infty$ , under suitable hypothesis (for instance, decay conditions on  $\alpha_{\#}\text{Leb}_{[0,1]}$ ).

For a given  $\alpha$ , we will often refer to the spectral risk measure  $\mathcal{R}_\alpha$  as the  $\alpha$ -risk, and  $\alpha$  as the the spectral function.

Spectral risk measures are pervasive in risk management and insurance. Indeed, it is well known that any coherent risk measure which is additive for comontonic random variables must be spectral (see for instance [13]).

We begin with the following variational characterization of spectral risk measures. This result is well known to experts (see, for example, [31] and [11], in which this fact is the basis for a multi-variate extension). We include a brief proof here for the convenience of the reader

**Lemma 12.** For any spectral risk measure and any probability measure  $\mu$  with  $\int_{\mathbb{R}} x d\mu(x) > -\infty$ ,

$$\mathcal{R}_\alpha(\mu) = \max_{\pi \in \Gamma(\alpha_{\#}\text{Leb}_{[0,1]}, \mu)} \int_{\mathbb{R} \times \mathbb{R}} xy d\pi(x, y), \quad (12)$$

where we have denoted  $\Gamma(\alpha_{\#}\text{Leb}_{[0,1]}, \mu)$  the space of probability measures on  $\mathbb{R}^2$  with marginals  $\alpha_{\#}\text{Leb}_{[0,1]}$  and  $\mu$ .

*Proof.* Let us re-write the right hand side of (12) as the following optimal transport problem between  $\mu_0 = \alpha_{\#}\text{Leb}_{[0,1]}$  and  $\mu_1 = \mu$

$$\max_{\gamma \in \Gamma(\mu_0, \mu_1)} \int_{[0,1] \times \mathbb{R}} s(x, y) d\pi(x, y),$$

where  $s(x, y) = xy$ . Then, since the surplus is strictly supermodular there exists a unique optimal  $\pi$  of the form  $\pi = (G_0, G_1)_{\#}\text{Leb}_{[0,1]}$ , where  $G_0(m) = F_{\mu_0}^{-1}(m) = F_{\alpha_{\#}\text{Leb}_{[0,1]}}^{-1}(m) = \alpha(m)$  and  $G_1(m) = F_{\mu}^{-1}(m)$ , by Theorem 7. The desired result follows.  $\square$

**Corollary 13.**  $\mathcal{R}_\alpha$  is concave on  $\mathcal{P}(\mathbb{R})$ .

*Proof.* Let  $\mu_0, \mu_1 \in \mathcal{P}(\mathbb{R})$  and let  $\pi_0, \pi_1$  be optimal for the problem (12). Then, with  $\mu_t = (1-t)\mu_0 + t\mu_1$  and  $\pi_t = (1-t)\pi_0 + t\pi_1$  one has  $\pi_t \in \Gamma(\alpha_{\#}\text{Leb}_{[0,1]}, \mu_t)$  so that

$$\begin{aligned} \mathcal{R}_\alpha(\mu_t) &\geq \int xy d\pi_t(x, y) \\ &= (1-t) \int xy d\pi_0(x, y) + t \int xy d\pi_1(x, y) \\ &= (1-t)\mathcal{R}_\alpha(\mu_0) + t\mathcal{R}_\alpha(\mu_1). \end{aligned}$$

□

Return now to the case in which there are  $d$  parameters which enter into the estimation of risk through the output function  $b$ , and we wish to evaluate the worst case scenario for the  $\alpha$ -risk  $\mathcal{R}_\alpha(b_{\#}\gamma)$  of  $b_{\#}\gamma$  among couplings  $\gamma \in \Gamma(\mu_1, \mu_2, \dots, \mu_d)$  of the marginal distributions  $\mu_i$ . That is, we want to maximize:

$$\max_{\gamma \in \Gamma(\mu_1, \mu_2, \dots, \mu_d)} \mathcal{R}_\alpha(b_{\#}\gamma). \quad (13)$$

**Remark 14** (Maximizing Expected Shortfall is optimal partial transport). *A special case of particular importance in risk management applications occurs when  $\alpha = \alpha_{m_0} := \frac{1}{m_0}\mathbf{1}_{[1-m_0, 1]}$ , in which case (11) is also known as the Expected Shortfall, or Conditional Value at Risk.*

*In this setting, the maximization problem (13) is actually equivalent to a well known variant of the optimal transport problem; indeed, it can be reformulated into*

$$\max_{\gamma \in \Gamma_{m_0}(\mu_1, \dots, \mu_d)} \frac{1}{m_0} \int b(x_1, \dots, x_d) d\gamma(x_1, \dots, x_d)$$

where  $\Gamma_{m_0}(\mu_1, \dots, \mu_d)$  denotes the set of non-negative measures  $\gamma$  on  $X_1 \times \dots \times X_d$  with total mass  $m_0$ , such that its  $i$ -marginal  $\gamma_i$  is dominated by  $\mu_i$  for each  $i$ ; that is  $\int_{X_i} \phi d\gamma_i \leq \int_{X_i} \phi d\mu_i$  for all non-negative test functions  $\phi \in \mathcal{C}^0(X_i)$ .

*This is known as the optimal partial transport problem when  $d = 2$  [6, 12], and the multi-marginal optimal partial transport problem [20] when  $d > 2$ .*

As we will show below, the problem is in fact equivalent to an ordinary multi-marginal optimal transport problem with an additional marginal; see Theorem 18 and Remark 19 below.

**Proposition 15.** *Let  $b : X_1 \times X_2 \times \dots \times X_d \rightarrow \mathbb{R}$  be an upper semi-continuous function bounded from below. Then the map  $\gamma \mapsto \mathcal{R}_\alpha(b_{\#}\gamma)$  is concave.*

*Proof.* Using (12) one gets

$$\begin{aligned} \mathcal{R}_\alpha(b_{\#}\gamma) &= \max_{\sigma \in \Gamma(\alpha_{\#}\text{Leb}_{[0,1]}, b_{\#}\gamma)} \int x_0 y d\sigma(x_0, y) \\ &= \max_{\pi \in \Gamma(\alpha_{\#}\text{Leb}_{[0,1]}, \gamma)} \int x_0 b(z) d\pi(x_0, z), \end{aligned}$$

where  $z = (x_1, \dots, x_d)$ . Then concavity of  $\gamma \mapsto \mathcal{R}_\alpha(b_{\#}\gamma)$  on  $\mathcal{P}(\mathbb{R}^d)$  follows as in Corollary 13.  $\square$

### 3 Equivalence between maximizing spectral risk measures and multi-marginal optimal transport

Our first main contribution is to show that the spectral risk maximization problem (13) is equivalent to the multi-marginal optimal transport problem (4) with  $X_0 = [\alpha(0), \alpha(1)] \subseteq \mathbb{R}$ ,  $\mu_0 = \alpha_{\#}\text{Leb}_{[0,1]}$ , the other  $X_i$  and  $\mu_i$  representing the domains and distributions of the underlying variables, respectively, and

$$s(x_0, x_1, \dots, x_d) = x_0 b(x_1, \dots, x_d). \quad (14)$$

**Lemma 16.** *Suppose that  $\pi \in \Gamma(\mu_0, \mu_1, \dots, \mu_d)$  and let  $\gamma$  be the  $(1, \dots, d)$ -marginal of  $\pi$ . Then, for  $s$  given by (14),*

$$\int_{\times_{i=0}^d X_i} s(x_0, x_1, \dots, x_d) d\pi(x_0, x_1, \dots, x_d) \leq \mathcal{R}_\alpha(b_{\#}\gamma)$$

Furthermore, we have equality if and only if the support of

$$\tau_\pi = \left( (x_0, x_1, x_2, \dots, x_d) \mapsto (x_0, b(x_1, x_2, \dots, x_d)) \right)_{\#} \pi \in \mathcal{P}(\mathbb{R}^2)$$

is monotone increasing.

*Proof.* We first note that

$$\begin{aligned} \int_{\times_{i=0}^d X_i} s(x_0, x_1, \dots, x_d) d\pi(x_0, \dots, x_d) &= \int_{\times_{i=0}^d X_i} x_0 b(x_1, \dots, x_d) d\pi(x_0, \dots, x_d) \\ &= \int_{\mathbb{R}^2} z x_0 d\tau_\pi(x_0, z) \\ &\leq \mathcal{R}_\alpha(b_\#(\gamma)), \end{aligned}$$

by Lemma 12. Furthermore, the uniqueness in Theorem 7 implies that the inequality is in fact an equality if and only if the coupling  $\tau_\pi$  has monotone increasing support.  $\square$

**Lemma 17.** *Given any measure  $\gamma \in \Gamma(\mu_1, \dots, \mu_d)$ , there exists a  $\pi \in \Gamma(\mu_0, \mu_1, \dots, \mu_d)$  whose  $(1, \dots, d)$ -marginal is  $\gamma$ , such that  $\left( (x_0, x_1, \dots, x_d) \mapsto (x_0, b(x_1, x_2, \dots, x_d)) \right)_\# \pi$  has monotone increasing support.*

*Proof.* Let  $\tau = (F_{\mu_0}^{-1}, F_{b_\# \gamma}^{-1})_\# \text{Leb}_{[0,1]}$  be the comonotonic coupling of  $\mu_0$  and  $b_\# \gamma$ . We disintegrate  $\tau(x_0, y) = \tau^y(x_0) \otimes (b_\# \gamma)(y)$  with respect to its second marginal, so that for any measurable function  $g(x_0, y)$  we have

$$\int_{\mathbb{R}^2} g(x_0, y) d\tau(x_0, y) = \int_{\mathbb{R}} \int_{\mathbb{R}} g(x_0, y) d\tau^y(x_0) d(b_\# \gamma)(y),$$

and  $\gamma(x_1, \dots, x_d) = \gamma^y(x_1, \dots, x_d) \otimes (b_\# \gamma)(y)$  with respect to  $b$ , so that for any measurable  $h(x_1, \dots, x_d)$  we have

$$\int_{X_1 \times \dots \times X_d} h(x_1, \dots, x_d) d\gamma(x_1, \dots, x_d) = \int_{\mathbb{R}} \int_{b^{-1}(y)} h(x_1, \dots, x_d) d\gamma^y(x_1, \dots, x_d) d(b_\# \gamma).$$

Any measure  $\pi(x_0, x_1, \dots, x_d) = \pi^y(x_0, x_1, \dots, x_d) \otimes (b_\# \gamma)(y)$  on  $X_0 \times (X_1 \times \dots \times X_d)$ , whose conditional probabilities  $\pi^y \in \Gamma(\tau^y, \gamma^y) \subseteq \mathcal{P}(X_0 \times X_1 \times \dots \times X_d)$  after disintegrating with respect to the mapping  $(x_0, \dots, x_d) \mapsto (x_0, b(x_1, \dots, x_d))$  are couplings of  $\tau^y$  and  $\gamma^y$  for  $b_\# \gamma$  almost every  $y$  then satisfies the requirement.  $\square$

We are now ready to establish the equivalence between the maximal  $\alpha$ -risk problem (13) and the multi-marginal optimal transport problem (4).



**Theorem 18.** *A probability measure  $\pi$  in  $\Gamma(\mu_0, \mu_1, \dots, \mu_d)$  is optimal for (4) with cost function (14) if and only if its  $(1, \dots, d)$ -marginal is optimal in (13), and  $\tau_\pi = \left( (x_0, x_1, x_2, \dots, x_d) \mapsto (x_0, b(x_1, x_2, \dots, x_d)) \right)_\# \pi$  has monotone increasing support.*

*Proof.* First, suppose that  $\pi$  is optimal in (4) and let  $\gamma$  be its  $(1, \dots, d)$ -marginal. Then for any other  $\tilde{\gamma} \in \Gamma(\mu_1, \dots, \mu_d)$ , construct  $\tilde{\pi} \in \Gamma(\mu_0, \mu_1, \dots, \mu_d)$  as in Lemma 17. We then have, by Lemma 16 and optimality of  $\pi$

$$\begin{aligned} \mathcal{R}_\alpha(b_\# \tilde{\gamma}) &= \int_{X_0 \times X_1 \times \dots \times X_d} s(x_0, \dots, x_d) d\tilde{\pi}(x_0, \dots, x_d) \\ &\leq \int_{X_0 \times X_1 \times \dots \times X_d} s(x_0, \dots, x_d) d\pi(x_0, \dots, x_d) \\ &\leq \mathcal{R}_\alpha(b_\# \gamma). \end{aligned}$$

This establishes that  $\gamma$  is optimal in (13). Furthermore, we must have equality throughout if we choose  $\tilde{\gamma} = \gamma$ . In particular, this implies that  $\int_{X_0 \times X_1 \times \dots \times X_d} s(x_0, \dots, x_d) d\pi(x_0, \dots, x_d) = \mathcal{R}_\alpha(b_\# \gamma)$ , which, by the second assertion in Lemma 16, implies that  $\tau_\pi$  has monotone increasing support.

Conversely, assume that  $\pi \in \Gamma(\mu_0, \mu_1, \dots, \mu_d)$ , such that its  $(1, \dots, d)$ -marginal  $\gamma$  is optimal in (13) and  $\pi$  couples  $\mu_0$  and  $b_\# \gamma$  monotonically. Let now  $\tilde{\pi}$  be an optimizer in (4) and  $\tilde{\gamma}$  its  $(1, \dots, d)$ -marginal. We then have, using Lemma 16,

$$\begin{aligned} \mathcal{R}_\alpha(b_\# \gamma) &= \int_{X_0 \times X_1 \times \dots \times X_d} s(x_0, x_1, \dots, x_d) d\pi(x_0, x_1, \dots, x_d) \\ &\leq \int_{X_0 \times X_1 \times \dots \times X_d} s(x_0, x_1, \dots, x_d) d\tilde{\pi}(x_0, x_1, \dots, x_d) \\ &\leq \mathcal{R}_\alpha(b_\# \tilde{\gamma}) \end{aligned}$$

Optimality of  $\gamma$  in (13) then implies that we must actually have equality; in particular  $\pi$  is optimal in (4) as desired.  $\square$

**Remark 19.** *Returning to the Expected Shortfall Case,  $\alpha = \alpha_{m_0} = \frac{1}{m_0} \mathbf{1}_{[1-m_0, 1]}$ , in view of Remark 14, Theorem 18 shows that the optimal partial transport problem for the surplus  $b(x_1, \dots, x_d)$  can be transformed into a multi-marginal transport problem for the surplus  $s(x_0, \dots, x_d) = x_0 b(x_1, \dots, x_d)$  and additional marginal  $\mu_0 = \alpha_\# \text{Leb}_{[0, 1]}$ . To the best of our knowledge, this equivalence*

between these two well studied mathematical problems has not been observed before in the optimal transport literature.

In addition, this perspective, together with results in [6, 12, 20] allows us to immediately identify conditions under which the active (that is, the part that couples to  $\alpha > 0$ ) part of the optimal  $\gamma$  in (13) is uniquely determined and concentrates on a graph over  $x_1$ . Furthermore, algorithms to compute the solution are readily available [3, 17].

These questions will be revisited for more general spectral risk measures later on.

We close this section with the following immediate consequence of Theorem 18 and a standard existence result in optimal transport theory.

**Corollary 20.** *Suppose that  $b$  is bounded above and upper-semicontinuous on  $X_1 \times X_2 \times \dots \times X_d$ . Then there exists a maximizing  $\gamma$  in (13).*

*Proof.* Note that the boundedness of  $\alpha$  implies that  $s$  given by (14) is bounded above and upper-semicontinuous on  $X_0 \times X_1 \times X_2 \times \dots \times X_d$ . The existence of an optimal  $\pi$  in (23) is then well known; see for example, Theorem 1.7 in [33]. The  $(1, \dots, d)$ -marginal of  $\pi$  then maximizes (13) by Theorem 18.  $\square$

## 4 Solutions for one-dimensional assets and compatible payouts

We now turn our attention to the structure of maximizers in (13) when the underlying variables are one dimensional.

Assume that each  $x_i \in \mathbb{R}$  and each  $\mu_i$  is supported on an interval,  $X_i = [\underline{x}_i, \bar{x}_i]$ , for all  $i = 1, \dots, d$ . We note that in our setting, the first marginal,  $\mu_0$  is supported always on the interval  $[\underline{x}_0, \bar{x}_0] := [\alpha(0), \alpha(1)]$  with  $\alpha(0) \geq 0$ .

**Lemma 21.** *Suppose that  $b$  is compatible and monotone increasing in each  $x_i \in S_+$  and monotone decreasing for each  $x_i \in S_-$ . Then the  $s$ -comonotone coupling (7) is optimal for (4) with surplus function given by (14). The maximal value is*

$$\int_0^1 \alpha(m) b(G_1(m), G_2(m), \dots, G_d(m)) dm. \quad (15)$$

Furthermore, if in addition the monotonicity of  $b$  with respect to each argument is strict, the solution is unique on the support of  $\alpha$ ; that is, any

other solution  $\bar{\pi}$  coincides with  $\pi = (G_\alpha, G_1, \dots, G_d)_{\#} \text{Leb}_{[0,1]}$  on  $(0, \alpha(1)] \times [\underline{x}_1, \bar{x}_1] \times \dots \times [\underline{x}_d, \bar{x}_d]$ .

Clearly full uniqueness of the optimal  $\pi$  cannot hold if  $\alpha = 0$  on a set of positive measure, or, equivalently,  $\mu_0(\{0\}) > 0$ ; in this case, we can rearrange the part of the  $(1, \dots, d)$ -marginal  $\gamma$  of  $\pi$  that couples to  $\alpha = 0$  in any way without affecting the value of  $\int s d\pi$ . The preceding Lemma identifies conditions under which this is the only source of non-uniqueness, so that the optimal  $\pi$  is uniquely determined on  $(0, \alpha(1)] \times [\underline{x}_1, \bar{x}_1] \times \dots \times [\underline{x}_d, \bar{x}_d]$ .

Note also that we are able to obtain uniqueness off the set where  $\alpha = 0$  despite the fact that the surplus function may not be strongly compatible. This is done by exploiting an idea developed in [27] (see in particular Proposition 4.1 there); namely that one only needs strongly compatible interactions between certain key pairs of variables (rather than all of them).

*Proof.* The conditions on  $b$  make the function  $s$  compatible, and so Proposition 9 implies the optimality of the  $s$ -comonotone  $\pi$ . Note that since  $\alpha$  is in fact the quantile function for the first marginal  $\mu_0 = \alpha_{\#} \text{Leb}_{[0,1]}$ , (15) is exactly  $\int s d\pi$ .

Now, to prove the last assertion, we will take advantage of the fact that compatibility is equivalent to supermodularity up to a change of variables (recall Remark 4). After such a change of variables, our assumptions become that  $b$  is supermodular, and strictly increasing in each argument, and each  $G_i = F_{\mu_i}^{-1}$  is the quantile map of  $\mu_i$ .

It will suffice to prove the following claim: for any optimizer  $\bar{\pi}$ , if  $(x_0, x_1, \dots, x_d)$  and  $(\tilde{x}_0, \tilde{x}_1, \dots, \tilde{x}_d)$  are both in the support of  $\bar{\pi}$  and  $x_0 > \tilde{x}_0 \geq 0$ , then  $x_i \geq \tilde{x}_i$  for each  $i$ . To see that this is sufficient, note that applying the claim for  $\tilde{x}_0 = 0$  implies that the lowest  $a$  portion of each mass,  $\{x_i < G_i(a)\}$  must pair with  $x_0 = 0$ , where  $a = \mu_0(\{0\})$ . Applying the claim with  $\tilde{x}_0 > 0$  then implies that the support of  $\bar{\pi}$  on  $(0, \alpha(1)] \times [\underline{x}_1, \bar{x}_1] \times \dots \times [\underline{x}_d, \bar{x}_d]$  must be monotone increasing, which immediately implies the desired result.

To see the claim, we apply the  $s$ -monotonicity property found, in, for example [25, 7]. We define  $x_i^+ = \max\{x_i, \tilde{x}_i\}$  and  $x_i^- = \min\{x_i, \tilde{x}_i\}$ . We then have

$$s(x_0^+, x_1^+, \dots, x_d^+) + s(x_0^-, x_1^-, \dots, x_d^-) \leq s(x_0, x_1, \dots, x_d) + s(\tilde{x}_0, \tilde{x}_1, \dots, \tilde{x}_d). \quad (16)$$

Note that  $x_0^- = \tilde{x}_0$ . If  $\tilde{x}_0 = 0$ , then  $s(x_0^-, x_1^-, \dots, x_d^-) = s(\tilde{x}_0, \tilde{x}_1, \dots, \tilde{x}_d) = 0$ , so that (16) becomes (after dividing by  $x_0$ )

$$b(x_1^+, \dots, x_d^+) \leq b(x_1, \dots, x_d)$$

As each  $x_i^+ \geq x_i$  and  $b$  is strictly monotone in each coordinate, this can only happen if each  $x_i^+ = x_i$ , which is equivalent to the claim.

On the other hand, if  $\tilde{x}_0 > 0$ , we follow the approach in the proof of Proposition 4.1 in [27]. A straightforward calculation yields that

$$\begin{aligned} & s(x_0^+, x_1^+, \dots, x_d^+) + s(x_0^-, x_1^-, \dots, x_d^-) - s(x_0, x_1, \dots, x_d) - s(\tilde{x}_0, \tilde{x}_1, \dots, \tilde{x}_d) \\ &= \int_0^1 \int_0^1 \sum_{i \neq j=0}^d \frac{\partial^2 s(x(\theta, \phi))}{\partial x_i \partial x_j} (x_i^+ - x_i)(x_j - x_j^-) d\theta d\phi \end{aligned} \quad (17)$$

where

$$x(\theta, \phi) = \phi[\theta x^+ + (1 - \theta)x] + (1 - \phi)[\theta \tilde{x} + (1 - \theta)x^-].$$

Now, by supermodularity of  $s$ , each  $\frac{\partial^2 s(x(\theta, \phi))}{\partial x_i \partial x_j} \geq 0$ , and as each  $x_i^+ - x_i \geq 0$  and  $x_j - x_j^- \geq 0$ , we have that (17) is greater than or equal to 0; by (16), it must then be equal to 0, which then implies that each individual term  $\frac{\partial^2 s(x(\theta, \phi))}{\partial x_i \partial x_j} (x_i^+ - x_i)(x_j - x_j^-)$  in the sum must be 0. Now, strict monotonicity of  $b$  with respect to each  $x_i$  implies that when we take  $j = 0$ , we have  $\frac{\partial^2 s(x(\theta, \phi))}{\partial x_i \partial x_0} = \frac{\partial b(x(\theta, \phi))}{\partial x_i} > 0$ , which, as  $x_0 > \tilde{x}_0 = x_0^-$  by assumption, implies that each  $x_i^+ = x_i$ . This completes the proof of the claim, and, consequently, the Lemma.  $\square$

Thanks to Theorem 18, the preceding result then easily yields the following characterization of solutions to (13).

**Theorem 22.** *Assume that  $b$  is compatible, monotone increasing in each  $x_i \in S_+$  and monotone decreasing in each  $x_i \in S_-$ . Then  $(G_1, \dots, G_d)_{\#} \text{Leb}_{[0,1]}$  maximizes (13) and the maximal value is given by (15). If in addition the monotonicity is strict, then letting  $a = \mu_0\{0\}$  any other maximizer  $\gamma$  must couple the regions defined by  $G_i([0, a])$ , and must coincide with the b-comonotone coupling  $(G_1, \dots, G_d)_{\#} \text{Leb}_{[0,1]}$  on  $G_1([a, 1]) \times G_2([a, 1]) \times \dots \times G_d([a, 1])$ .*

Note that in the special case when  $b$  is supermodular and increasing with respect to each argument, this asserts optimality of the comonotone coupling. Note that each region  $G_i([0, a])$  is of mass  $a$ ; it is an interval of the form  $[x_i, G_i(a)]$  if  $i \in S_+$  (corresponding to the lowest fraction of the mass) and of the form  $[G_i(a), \bar{x}_i]$  if  $i \in S_-$  (corresponding to the highest fraction of the mass).

**Example 23.** Consider  $b(x_1, \dots, x_d) = x_1 + \dots + x_d$ . This corresponds to the value of a portfolio composed of underlying assets  $x_1, \dots, x_d$ . It is clearly weakly supermodular and therefore compatible, and strictly monotone increasing in each coordinate, so the  $(1, \dots, d)$ -marginal of the  $\pi$  defined by (7) is optimal by Theorem 22.

Explicitly, the optimal value in (13) is:

$$\mathcal{R}_\alpha(b_{\#}\gamma) = \int_0^1 \alpha(m)(F_{\mu_1}^{-1}(m) + \dots + F_{\mu_d}^{-1}(m))dm = \sum_{i=1}^d \mathcal{R}_\alpha(\mu_i),$$

which corresponds to coupling “good” and “bad” events (i.e., with  $x_i$  small or large, respectively) together.

Note that this special case could alternatively be deduced by noting that spectral risk measures are sub-additive for all couplings, and additive for the comonotone coupling; see, for instance, [29] for the expected shortfall case (arguments for other  $\alpha$  are similar).

**Example 24.** If  $b(x_1, \dots, x_d) = -\sum_{1 \leq i < j \leq d} (x_i - x_j)^2$  and if  $\alpha \equiv 1$ , problem (13) is equivalent to the computation of Wasserstein barycenters [1], while with  $\alpha = \alpha_{m_0}$  we get the partial Wasserstein barycenter problem [20]. In both cases, since  $b$  is supermodular the answer can be calculated explicitly by Theorem 22.

**Example 25.** For  $\alpha = 1$ , it is well known that optimal transport for the surplus function in the preceding example is equivalent to the same problem with  $b(x_1, \dots, x_d) = |\sum_{i=1}^d x_i|^2$  (see, for example, [14]). Since the mean  $M := \int_{X_1 \times \dots \times X_d} \sum_{i=1}^d x_i d\gamma(x_1, \dots, x_d) = \sum_{i=1}^d \int_{X_i} x_i d\mu_i(x_i)$  of the sum  $z = \sum_{i=1}^d x_i$  is completely determined by the marginals  $\mu_i$  for any  $\gamma \in \Gamma(\mu_1, \dots, \mu_d)$ , (4) is equivalent to maximizing the variance of the sum  $z$  over  $\gamma \in \Gamma(\mu_1, \dots, \mu_d)$ . Interpreting this sum as the losses of a portfolio comprised of assets with losses  $x_i$ , we note the variance is sometimes used in this setting as a measure of risk. One drawback to this approach is its

*symmetry; variance punishes good outcomes (when the sum  $\sum_{i=1}^d x_i$  is below the mean) as well as bad ones (when the sum  $\sum_{i=1}^d x_i$  is above the mean).*

*Choosing  $\alpha = \alpha_{m_0}$  in (11) on the other hand measures only downside risk beyond a particular level. Since the quadratic surplus is clearly supermodular, Theorem 22 implies that the worse case scenario can be computed explicitly.*

*More generally, surplus functions of the form  $b(x_1, \dots, x_d) = H(\sum_{i=1}^d x_i)$  for a convex function  $H$  are supermodular so Theorem 22 applies in this setting as well. Note that the case  $H(z) = \max(z - K, 0)$  reflects the payout of a basket call option on the underlying assets.*

Our last example requires more extensive explanation and so we devote a separate subsection to it.

## 4.1 Sensitivity analysis and maximal river flow

In this section we briefly describe a recurring example from [18], used throughout that paper to illustrate issues in sensitivity analysis. In that setting, one wants to understand the influence of the dependence structure (among other factors) between several contributing inputs on an output behavior.

We consider a simple model which involves the height of river at risk of flooding and compares it to the height of a dyke which protects industrial facilities. The maximal annual overflow  $S$  of a river is modelled by

$$S = Z_\nu + \left( \frac{Q}{BK_s \sqrt{\frac{Z_m - Z_\nu}{L}}} \right)^{0.6} - H_d - C_b,$$

where  $\left( \frac{Q}{BK_s \sqrt{\frac{Z_m - Z_\nu}{L}}} \right)^{0.6}$  is the maximal annual height of the river and the variables  $Q, K_s, Z_\nu, Z_m, H_d, C_b, L, B$  are physical quantities whose values are modelled as random variables due to their variation in time and space, measurement inaccuracies, or uncertainty of their true values. Their individual distributions are modelled in [18] (see Table 1 on p.4), and are all absolutely continuous with respect to Lebesgue measure. The  $\alpha$ -risk (11), where the  $x_i$  are the variables  $Q, K_s, Z_\nu, Z_m, H_d, C_b, L, B$  and  $b = S$  is the overflow, quantifies the risk of overflow. In [18], the variables were assumed to be independent, although other dependence structures are certainly possible; (13) asks what is the maximal risk over all possible dependence structures.

Notice then that the surplus function  $b$  is compatible and satisfies the strict monotonicity with respect to each variable required in Theorem 22. The  $(1, \dots, d)$  marginal  $\gamma$  of the  $s$ -comonotone solution  $\pi$  defined in (7) is optimal, and the unique optimizer on the support of  $\alpha$ ,  $\{\alpha > 0\}$ . In this case, the  $G_i$  corresponding to  $Z_\nu, Q$  and  $L$  are monotone increasing, while the  $G_i$  corresponding to the other variables are decreasing.

They can be retrieved explicitly by computing the quantile functions associated to each marginals as described above.

## 5 Higher dimensional assets

There is an analogous theory of multi-marginal optimal transport when the underlying variables lie in more general spaces. Although in these cases it is generally not possible to derive explicit solutions as we did above, it is possible to prove analogous structural properties of optimal couplings, namely, that solutions are of Monge type (that is, concentrated on graphs over  $x_1$ ), for certain surplus functions  $s$  (see [14] for an early result in this direction, for a particular  $s$ , and [19] and [23] for general, sufficient conditions on  $s$ ).

We present one such result below, when each  $X_i \subseteq \mathbb{R}^n$ , to illustrate how the theory can be adapted to the present setting. For simplicity, we restrict our attention to the  $d = 2$  case. The conditions we impose on  $b$  are closely related to the local differential condition found in [23]; however, the proof in [23] was restricted to the case when each marginal was supported on a space of the same dimension, whereas here our first marginal  $\mu_0$ , corresponding to  $x_0 = \alpha$ , has one dimensional support while the other marginals are supported in  $\mathbb{R}^n$ . We modify that proof to fit the present setting here. Similar results can be established for larger  $d$ , using similar arguments. However, the conditions imposed on  $b$  become more stringent for larger  $d$ , and more complicated to state.

In what follows, we will assume that  $b$  is  $\mathcal{C}^2$ ;  $D_{x_i} b(x_1, x_2) = (\frac{\partial b}{\partial x_i^1}, \frac{\partial b}{\partial x_i^2}, \dots, \frac{\partial b}{\partial x_i^n})$  represents the gradient of the function  $b : \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}$  with respect to the variable  $x_i \in \mathbb{R}^n$ , for  $i = 1, 2$ . Similarly,  $D_{x_1 x_2}^2 b(x_1, x_2) = (\frac{\partial^2 b}{\partial x_1^r \partial x_2^l})_{rl}$  is the  $n$  by  $n$  matrix of mixed second order derivatives (each entry is a derivative with respect to one coordinate from each of  $x_1$  and  $x_2$ ).

**Theorem 26.** *Suppose that  $d = 2$ , that the domains  $X_1, X_2 \subseteq \mathbb{R}^n$  are compact and that  $\mu_1$  is absolutely continuous with respect to Lebesgue measure. Assume that  $x_2 \mapsto D_{x_1}b(x_1, x_2)$  is injective for each fixed  $x_1 \in X_1$ , and that for each  $(x_1, x_2) \in X_1 \times X_2$ ,  $\det(D_{x_1x_2}^2b(x_1, x_2)) \neq 0$  and*

$$D_{x_2}b(x_1, x_2) \cdot [D_{x_1x_2}^2b(x_1, x_2)]^{-1}D_{x_1}b(x_1, x_2) > 0. \quad (18)$$

*Then the part of the solution to (13) away from  $\alpha = 0$  concentrates on the graph of a function over  $x_1$ . Furthermore, if  $|\{\alpha = 0\}| = 0$ , the solution is unique.*

**Remark 27.** *The injectivity of  $x_2 \mapsto D_{x_1}b(x_1, x_2)$  is known as the twist condition in the optimal transport literature; it is well known that, together with the absolute continuity of  $\mu_1$ , it guarantees the Monge structure of the solution to the two marginal optimal transport problem with surplus  $b$  (see, for example, [33]). The invertibility of  $D_{x_1x_2}^2b(x_1, x_2)$  is frequently referred to as non-degeneracy, and can be seen as a linearized version of the twist. Here we require additional hypotheses as we are dealing with the more sophisticated 3 marginal problem with surplus  $s(\alpha, x_1, x_2) = \alpha b(x_1, x_2)$ , which is equivalent to the spectral risk maximization (13) by Theorem 18.*

*Proof.* By Theorem 18, we can consider the multi-marginal optimal transport problem (4) rather than (13). Let  $\pi$  solve (4) and  $u_0, u_1, u_2$  the optimal solutions to the dual problem (9). Then  $u_1$  is differentiable  $\mu_1$  almost everywhere, by a standard argument, originally found in [21]. Fix an  $x_1$  where  $u_1$  is differentiable. To show that the solution concentrates on a graph, we need to prove that there is only one  $x_2$  and one non-zero  $\alpha$  such that  $(\alpha, x_1, x_2)$  is in the support of  $\pi$ .

At such points, the inequality constraint in (9), together with the equality (10) on the support of the optimizer and the envelope theorem implies, whenever  $(\alpha, x_1, x_2)$  is in the support of the optimal  $\pi$

$$Du_1(x_1) = \alpha D_{x_1}b(x_1, x_2).$$

Note that for fixed  $x_1$ , the twist condition implies that for  $\alpha \neq 0$ , this uniquely defines  $x_2 = x_2(\alpha)$  as a function of  $\alpha$  (that is, for each  $\alpha$ , there is at most one  $x_2$  satisfying this equation). Furthermore, we can differentiate implicitly with respect to  $\alpha$  to obtain,

$$0 = D_{x_1}b(x_1, x_2(\alpha)) + D_{x_1x_2}^2b(x_1, x_2(\alpha))D_\alpha x_2(\alpha),$$



or

$$D_\alpha x_2(\alpha) = -[D_{x_1 x_2}^2 b(x_1, x_2(\alpha))]^{-1} D_{x_1} b(x_1, x_2(\alpha)). \quad (19)$$

Now, it is well known (see for example [14] or [23]) that the  $u_i$  can be taken to be  $s$ -conjugate. Explicitly, this means that  $u_0$ , the potential corresponding to  $x_0 = \alpha$ , can be written as

$$u_0(\alpha) = \sup_{x_1, x_2} [\alpha b(x_1, x_2) - u_1(x_1) - u_2(x_2)],$$

and is therefore convex, as a supremum of affine functions. If  $(\alpha, x_1, x_2)$  is in the support of the optimal  $\pi$ , by the optimality condition presented in section 2, we must have that  $x_1, x_2$  attains the supremum above, so that  $b(x_1, x_2)$  lies in the subdifferential of  $u_0$  at  $\alpha$ :

$$b(x_1, x_2) \in \partial u_0(\alpha).$$

But we also have  $x_2 = x_2(\alpha)$ , so that

$$b(x_1, x_2(\alpha)) \in \partial u_0(\alpha). \quad (20)$$

We claim that this can hold for at most one value of  $\alpha$ . Indeed, differentiating  $b(x_1, x_2(\alpha))$  with respect to  $\alpha$  and using (19) yields

$$\begin{aligned} D_\alpha b(x_1, x_2(\alpha)) &= D_{x_2} b(x_1, x_2(\alpha)) \cdot D_\alpha x_2(\alpha) \\ &= -D_{x_2} b(x_1, x_2(\alpha)) \cdot [D_{x_1 x_2}^2 b(x_1, x_2(\alpha))]^{-1} D_{x_1} b(x_1, x_2(\alpha)). \end{aligned}$$

Under the given assumption, this quantity is negative, and therefore the left hand side in (20) is a strictly decreasing function of  $\alpha$ ; it can therefore intersect the subdifferential  $\partial u_0(\alpha)$ , an increasing set valued function of  $\alpha$ , at most once. The points  $(\alpha, x_2(\alpha))$ , where  $\alpha$  is the point at which this intersection occurs, are then the only points such that  $(\alpha, x_1, x_2)$  are in the support of  $\pi$ , which establishes that the solution  $\gamma$  to (13) concentrates on a graph.

Uniqueness then follows by a very standard argument; if  $\pi_0$  and  $\pi_1$  are both solutions, by linearity of the functional  $\gamma_{1/2} = \frac{1}{2}[\gamma_0 + \gamma_1]$  is also a solution. The argument above implies that the supports of  $\gamma_0$  and  $\gamma_1$  concentrate on graphs  $T_0$  and  $T_1$  over  $x_1$ . The support of  $\gamma_{1/2}$  then concentrates on the union of the graphs of  $T_0$  and  $T_1$ . However, since  $\gamma_{1/2}$  is also a solution, the argument above implies that this set should concentrate on a single graph. This is not possible unless  $T_0 = T_1$ ,  $\mu_1$  almost everywhere, in which case  $\gamma_0$  and  $\gamma_1$  coincide, establishing the uniqueness assertion.  $\square$

There are of course output functions  $b$  for which the hypotheses in the preceding result do not hold, and in these cases solutions are often not of Monge type. Alternatively, one can estimate the dimension of the support of the optimal  $\pi$  in (4) using the signature of the off diagonal part of the Hessian of  $s(\alpha, x_1, \dots, x_d) = \alpha b(x_1, \dots, x_d)$ , as in [22] and [24]. We illustrate this below with the following result in the  $d = 2$  case, which is an immediate consequence of Theorem 3.1.3 and Lemma 3.3.2 in [22].

**Proposition 28.** *Assume  $X_1, X_2 \subseteq \mathbb{R}^n$  and choose any point  $(\alpha, x_1, x_2) \in (0, \infty) \times X_1 \times X_2$  where  $D_{x_1 x_2}^2 b(x_1, x_2)$  is non-singular. Let  $\pi$  be optimal in (4). Then in a neighbourhood of  $(\alpha, x_1, x_2)$  the support of  $\pi$  is contained in a Lipschitz submanifold of dimension at most  $n + 1$ . Furthermore, if (18) holds, the dimension is at most  $n$ .*

For larger  $d$ , if each  $X_i \subseteq \mathbb{R}^{n_i}$  lies in a space of dimension  $n_i$ , the results in [22] and [24] imply that the support of  $\pi$  will again be contained in a Lipschitz submanifold of dimension  $m$ , where  $m$  depends on the mixed second derivatives of  $b$  in an intricate way. Generically, if each  $n_i = n$  is the same,  $m$  will lie between  $n$  and  $(d - 1)n + 1$ . Of course, the dimension of the support of the solution  $\gamma$  to (13) will be no larger than  $m$ , as by Theorem 18  $\gamma$  is the  $(1, \dots, d)$ -marginal of  $\pi$ .

## 6 Stability

Among other results, in [15] the authors show stability of the optimal value  $\mathcal{R}_\alpha(b_\# \gamma)$  and measure  $\gamma$  with respect to weak convergence of the marginals and  $L^1$  convergence of the payoff function  $b$ . Below, we show that similar results (under slightly different assumptions – see Remark 30) can be easily deduced by combining Theorem 18 with optimal transport theory.

In addition, we further leverage the connection to optimal transport to improve the stability with respect to the marginals when the underlying assets are one dimensional and  $b$  is compatible; in this case, we show that under appropriate conditions,  $\mathcal{R}_\alpha$  in fact exhibits Lipschitz dependence on the marginals.

**Proposition 29.** *Suppose that for each  $i$   $\{\mu_i^k\}$  is a sequence of probability measures in  $\mathcal{P}(X_i)$  converging weakly to  $\mu_i$  as  $k \rightarrow \infty$  and  $\{b^k\}$  is a sequence of continuous output functions on  $X_1 \times X_2 \times \dots \times X_d$  converging uniformly*

to a continuous  $b$ , such that  $b$  is bounded above, that is  $b \leq C$ , and bounded below by a sum of integrable functions, that is,  $b(x_1, \dots, x_d) \geq \sum_{i=1}^d u_i(x_i)$  with each  $u_i \in L^1(\mu_i)$ . Let  $\alpha^k$  be a sequence of spectral functions, uniformly bounded  $\alpha^k \leq M$ , such that the associated spectral measure  $\alpha^k_{\#} \text{Leb}_{[0,1]}$  converges weakly to  $\alpha_{\#} \text{Leb}_{[0,1]}$ , where  $\alpha$  is bounded. Then

$$\max_{\gamma \in \Gamma(\mu_1^k, \mu_2^k, \dots, \mu_d^k)} \mathcal{R}_{\alpha^k}(b_{\#}^k \gamma) \rightarrow \max_{\gamma \in \Gamma(\mu_1, \mu_2, \dots, \mu_d)} \mathcal{R}_{\alpha}(b_{\#} \gamma).$$

Furthermore, if  $\gamma^k$  maximizes  $\mathcal{R}_{\alpha^k}(b_{\#}^k \gamma)$  over  $\Gamma(\mu_1^k, \mu_2^k, \dots, \mu_d^k)$ , then the weak limit  $\gamma$  of any weakly convergent subsequence of the  $\gamma^k$  maximizes  $\mathcal{R}_{\alpha}(b_{\#} \gamma)$  over  $\Gamma(\mu_1, \mu_2, \dots, \mu_d)$ .

**Remark 30.** This result should be compared with Proposition 7 in [15]; in that setting, stronger assumptions (uniform Holder continuity) on the  $b^k$  are required, although a weaker notion of convergence is imposed on them. In addition, we also obtain stability with respect to variations in the spectral function  $\alpha$ .

*Proof.* The proof follows almost immediately from Theorem 18 and a corresponding stability result in optimal transport; namely, [35, Theorem 5.20]. Note that as we can take  $X_0 = [0, M]$  to be bounded, uniform convergence of the  $b^k$  implies uniform convergence of the corresponding  $s^k$  defined by (14).

The only issue to note is that that [35, Theorem 5.20] applies only to the two marginal problem. However, exactly the same argument as is used there applies to the multi-marginal case to show that the multi-marginal version of the  $s^k$ -cyclic monotonicity property satisfied by the support of each  $\pi^k$  implies that the support of the limit measure  $\pi$  is  $s$ -cyclically monotone. This then implies that  $\pi$  is optimal in (4) by [16, Theorem 12].

□

We now specialize to the case when the  $x_i$  are one dimensional. Our main result in this direction requires the following lemma. It is in fact a special case of [28, Corollary 11]; we provide the short proof for the convenience of the reader.

**Lemma 31.** *If  $\alpha \leq M$ , then  $\mu \in \mathcal{P}(\mathbb{R}) \mapsto \mathcal{R}_{\alpha}(\mu)$  is  $M$ -Lipschitz for the  $p$ -Wasserstein distance,*

$$|\mathcal{R}_{\alpha}(\mu) - \mathcal{R}_{\alpha}(\nu)| \leq MW_p(\mu, \nu). \quad (21)$$

*Proof.* On  $\mathbb{R}$ , the  $p$ -Wasserstein distance is given by the  $L^p$  distance between the inverse cumulative distribution functions  $W_p(\mu, \nu) = \|F_\mu^{-1} - F_\nu^{-1}\|_{L^p([0,1])}$ . The statement follows:

$$\begin{aligned} \mathcal{R}_\alpha(\mu) - \mathcal{R}_\alpha(\nu) &= \int (F_\mu^{-1}(m) - F_\nu^{-1}(m))\alpha(m)dm \\ &\leq \|F_\mu^{-1} - F_\nu^{-1}\|_{L^p([0,1])} \|\alpha\|_\infty. \end{aligned}$$

□

**Proposition 32.** *Assume that each  $X_i \subset \mathbb{R}$  and that the output function  $b$  is weakly compatible, monotone increasing in each  $x_i \in S_+$  and monotone decreasing in each  $x_i \in S_-$ . If in addition  $b$  is  $K$ -Lipschitz with respect to  $\|\cdot\|_p$  on  $\mathbb{R}^d$  and  $\alpha$  is bounded by  $M$ , then:*

$$\left| \max_{\gamma \in \Gamma(\mu_1, \dots, \mu_d)} \mathcal{R}_\alpha(b_{\#}\gamma) - \max_{\gamma \in \Gamma(\nu_1, \dots, \nu_d)} \mathcal{R}_\alpha(b_{\#}\gamma) \right| \leq MK \left( \sum_i W_p(\mu_i, \nu_i) \right)^{1/p} \quad (22)$$

*Proof.* By Theorem 22, the  $(1, \dots, d)$ - marginals  $\gamma = (G_1, \dots, G_d)_{\#}\text{Leb}_{[0,1]}$  and  $\tilde{\gamma} = (\tilde{G}_1, \dots, \tilde{G}_d)_{\#}\text{Leb}_{[0,1]}$  of the  $s$ -monotone couplings  $\pi$  and  $\tilde{\pi}$  of the  $\mu_i$  and  $\nu_i$ , respectively, achieve the suprema of  $\mathcal{R}_\alpha$ . Therefore, we get that  $(G_1, \dots, G_d, \tilde{G}_1, \dots, \tilde{G}_d)_{\#}\text{Leb}_{[0,1]} \in \Gamma(\gamma, \tilde{\gamma}) \subset \mathcal{P}(\mathbb{R}^{2d})$ , and its projection onto each  $X_i \times X_i$  is  $(G_i, \tilde{G}_i)_{\#}\text{Leb}_{[0,1]}$ , which is the optimal coupling between  $\mu_i$  and  $\nu_i$  for the cost function  $|x_i - \tilde{x}_i|^p$  in the definition of the  $p$ -Wasserstein distance. Therefore, setting  $G = (G_1, \dots, G_d)$  and  $\tilde{G} = (\tilde{G}_1, \dots, \tilde{G}_d)$ ,

$$\begin{aligned} W_p^p(b_{\#}\gamma, b_{\#}\tilde{\gamma}) &\leq \int_0^1 (b(G(t)) - b(\tilde{G}(t)))^p dt \\ &\leq K^p \int \|G(t) - \tilde{G}(t)\|_p^p dt \\ &= K^p \int \sum_i |G_i(t) - \tilde{G}_i(t)|^p dt \\ &= K^p \sum_i W_p^p(\mu_i, \nu_i) \end{aligned}$$

One concludes using Lemma 31 on the Lipschitz continuity of  $\mathcal{R}_\alpha$ . □

## 7 Multidimensional measures of risk

We consider the framework in [31] in which risk is measured in a multi-dimensional way. Instead of a single output variable, we now have several, depending on the same underlying random variables; this corresponds to a vector valued output function  $b : X_1 \times \dots \times X_d \rightarrow \mathbb{R}^n$ . A natural form of multi-variate risk measures on the distribution  $b_{\#}\gamma$  of output variables is then the maximal correlation measure from [31], which is defined by

$$\mathcal{R}_\nu(b_{\#}\gamma) = \max_{\eta \in \Gamma(b_{\#}\gamma, \nu)} \int_{\mathbb{R}^n \times \mathbb{R}^n} z \cdot y \, d\eta$$

for some probability measure  $\nu \in \mathcal{P}(\mathbb{R}^n)$ . We note that it was proven in [11] that any strongly coherent multi-variate risk measure takes this form. We consider the problem of maximizing  $\mathcal{R}_\nu(b_{\#}\gamma)$  over all  $\gamma \in \Gamma(\mu_1, \dots, \mu_d)$ , where the  $\mu_i$  as before represent the distributions of the underlying variables. Exactly as in Theorem 18, one can show that this problem is equivalent to the multi-marginal problem

$$\max_{\pi \in \Gamma(\nu, \mu_1, \dots, \mu_d)} \int b(x_1, \dots, x_d) \cdot y \, d\pi. \quad (23)$$

This problem is more challenging than the case of a scalar valued  $b$ ; nonetheless, we are able to obtain the following results in particular cases.

**Proposition 33.** *Suppose that the underlying variables  $x_i$  are one dimensional. Assume that  $\nu$  is concentrated on a smooth curve, that is  $\nu = f_{\#}\text{Leb}_{[0,1]}$  with  $f : [0, 1] \rightarrow \mathbb{R}^d$ , where each component of  $f$  is positive and monotone increasing, and each component  $b_j$  of  $b = (b_1, b_2, \dots, b_n)$  is supermodular and monotone increasing in each  $x_i$ . Then  $\left( (m, x_1, \dots, x_d) \mapsto (f(m), x_1, \dots, x_d) \right)_{\#} \pi$  is optimal in (23), where  $\pi = (\text{id}, F_{\mu_1}^{-1}, F_{\mu_2}^{-1}, \dots, F_{\mu_d}^{-1})_{\#}\text{Leb}_{[0,1]}$  is the comonotone coupling of  $\mu_0, \mu_1, \dots, \mu_d$ , where  $\mu_0 = \text{Leb}_{[0,1]}$ .*

*Proof.* We rewrite the problem (23) above as

$$\max_{\pi \in \Gamma(\nu, \mu_1, \dots, \mu_d)} \int b(x_1, \dots, x_d) \cdot y \, d\pi = \max_{\pi \in \Gamma(\text{Leb}_{[0,1]}, \mu_1, \dots, \mu_d)} \int b(x_1, \dots, x_d) \cdot f(m) \, d\pi.$$

The objective function  $b(x_1, \dots, x_d) \cdot f(m) = \sum_{j=1}^d b_j(x_1, \dots, x_d) f_j(m)$  is now a supermodular function of the one dimensional variables  $m, x_1, x_2, \dots, x_d$ , and so Proposition 9 implies the desired result.  $\square$

For more general, diffuse  $\nu$ , we are able to prove that the solution is of Monge form and unique, provided that  $b$  is invertible.

**Proposition 34.** *Assume that  $\nu$  is absolutely continuous with respect to  $n$  dimensional Lebesgue measure and  $b : X_1 \times \dots \times X_d \rightarrow \mathbb{R}^n$  is invertible. Then there exists a unique solution to (23). Furthermore, it concentrates on a graph over  $y$ .*

Note that the invertibility assumption on  $b$  implies that the sum of the dimensions of the  $X_i$  must be less than or equal to the dimension  $n$  of  $\nu$

*Proof.* Noting that the gradient of the surplus function in (23) with respect to  $y$  is  $b$ , the invertibility condition on  $b$  is the famous twist condition in optimal transport, and so this result follows by a minor refinement of a standard argument which we produce below.

Letting  $u_0(y), u_1(x_1), \dots, u_n(x_n)$  be the optimal solutions to the dual problem, we note that as  $\sum_{i=0}^d u_i(x_i) \geq b(x_1, \dots, x_d) \cdot y$ , with equality by (10) in the support of any optimizer  $\pi$ , the envelope theorem implies,

$$Du_0(y) = D_y(y \cdot b(x_1, \dots, x_d)) = b(x_1, \dots, x_d) \quad (24)$$

wherever  $u_0$  is differentiable, at points  $(y, x_1, \dots, x_d)$  in the support of  $\pi$ . A now standard argument, found in [21], implies that  $u_0$  is Lipschitz and therefore differentiable Lebesgue almost everywhere by Rademacher's theorem; the differentiability holds  $\nu$  almost everywhere by the absolute continuity of  $\nu$ . Therefore, (24) holds on a set of full  $\pi$  measure. Invertibility of  $b$  then means that this equation becomes  $(x_1, \dots, x_d) = b^{-1}(Du_0(y))$ . Setting  $T(y) = b^{-1}(Du_0(y))$ , this means that  $\pi$  concentrates on the graph of  $T : \mathbb{R}^n \rightarrow \mathbb{R}^d$ , as desired.

Uniqueness then follows exactly as in the proof of Theorem 26.  $\square$

## References

- [1] M. Agueh and G. Carlier. Barycenters in the Wasserstein space. *SIAM J. on Mathematical Analysis*, 43(2):904–924, 2011.
- [2] Jean-David Benamou. Optimal transportation, modelling and numerical simulation. *Acta Numer.*, 30:249–325, 2021.

- [3] Jean-David Benamou, Guillaume Carlier, Marco Cuturi, Luca Nenna, and Gabriel Peyré. Iterative Bregman projections for regularized transportation problems. *SIAM J. Sci. Comput.*, 37(2):A1111–A1138, 2015.
- [4] Yann Brenier. The least action principle and the related concept of generalized flows for incompressible perfect fluids. *Journal of the American Mathematical Society*, 2(2):225–255, 1989.
- [5] Giuseppe Buttazzo, Luigi De Pascale, and Paola Gori-Giorgi. Optimal-transport formulation of electronic density-functional theory. *Physical Review A*, 85(6):062502, 2012.
- [6] Luis A. Caffarelli and Robert J. McCann. Free boundaries in optimal transport and Monge-Ampère obstacle problems. *Ann. of Math. (2)*, 171(2):673–730, 2010.
- [7] Guillaume Carlier. On a class of multidimensional optimal transportation problems. *Journal of convex analysis*, 10(2):517–530, 2003.
- [8] Guillaume Carlier and Bruno Nazaret. Optimal transportation for the determinant. *ESAIM Control Optim. Calc. Var.*, 14(4):678–698, 2008.
- [9] Codina Cotar, Gero Friesecke, and Claudia Klüppelberg. Density functional theory and optimal transportation with coulomb cost. *Communications on Pure and Applied Mathematics*, 66(4):548–599, 2013.
- [10] I. Ekeland. An optimal matching problem. *ESAIM Control. Optim. Calc. Var.*, 11:57–71, 2005.
- [11] Ivar Ekeland, Alfred Galichon, and Marc Henry. Comonotonic measures of multivariate risks. *Mathematical Finance: An International Journal of Mathematics, Statistics and Financial Economics*, 22(1):109–132, 2012.
- [12] Alessio Figalli. The optimal partial transport problem. *Arch. Ration. Mech. Anal.*, 195(2):533–560, 2010.
- [13] Hans Föllmer and Alexander Schied. *Stochastic finance*, volume 27 of *De Gruyter Studies in Mathematics*. Walter de Gruyter & Co., Berlin, 2002. An introduction in discrete time.

- [14] Wilfrid Gangbo and Andrzej Swiech. Optimal maps for the multidimensional monge-kantorovich problem. *Communications on pure and applied mathematics*, 51(1):23–45, 1998.
- [15] Mario Ghossoub, Jesse Hall, and David Saunders. Maximum spectral measures of risk with given risk factor marginal distributions. *To appear in Mathematics of Operations Research*. Preprint available at *arXiv:2010.14673*, 2020.
- [16] Claus Griessler.  $C$ -cyclical monotonicity as a sufficient criterion for optimality in the multimarginal Monge-Kantorovich problem. *Proc. Amer. Math. Soc.*, 146(11):4735–4740, 2018.
- [17] Nouredine Igbida and Van Thanh Nguyen. Augmented Lagrangian method for optimal partial transportation. *IMA J. Numer. Anal.*, 38(1):156–183, 2018.
- [18] Bertrand Iooss and Paul Lemaître. A review on global sensitivity analysis methods. *Uncertainty management in simulation-optimization of complex systems*, pages 101–122, 2015.
- [19] Young-Heon Kim and Brendan Pass. A general condition for monge solutions in the multi-marginal optimal transport problem. *SIAM Journal on Mathematical Analysis*, 46(2):1538–1550, 2014.
- [20] Jun Kitagawa and Brendan Pass. The multi-marginal optimal partial transport problem. *arXiv preprint arXiv:1401.7255*, 2014.
- [21] Robert J. McCann. Polar factorization of maps on Riemannian manifolds. *Geom. Funct. Anal.*, 11(3):589–608, 2001.
- [22] Brendan Pass. *Structural results on optimal transportation plans*. PhD thesis, University of Toronto, 2011.
- [23] Brendan Pass. Uniqueness and monge solutions in the multimarginal optimal transportation problem. *SIAM Journal on Mathematical Analysis*, 43(6):2758–2775, 2011.
- [24] Brendan Pass. On the local structure of optimal measures in the multi-marginal optimal transportation problem. *Calculus of Variations and Partial Differential Equations*, 43(3-4):529–536, 2012.



- [25] Brendan Pass. Regularity of optimal transportation between spaces with different dimensions. *Math. Res. Lett.*, 19(2):291–307, 2012.
- [26] Brendan Pass. Multi-marginal optimal transport: theory and applications. *arXiv preprint arXiv:1406.0026*, 2014.
- [27] Brendan Pass and Adolfo Vargas-Jiménez. Monge solutions and uniqueness in multi-marginal optimal transport: weaker conditions on the cost, stronger conditions on the marginals. *arXiv preprint arXiv:2202.06783*, 2022.
- [28] Alois Pichler. Evaluations of risk measures for different probability measures. *SIAM J. Optim.*, 23(1):530–551, 2013.
- [29] Giovanni Puccetti. Sharp bounds on the expected shortfall for a sum of dependent random variables. *Statistics & Probability Letters*, 83(4):1227–1232, 2013.
- [30] Giovanni Puccetti and Ludger Rüschendorf. Sharp bounds for sums of dependent risks. *J. Appl. Probab.*, 50(1):42–53, 2013.
- [31] Ludger Rüschendorf. Law invariant convex risk measures for portfolio vectors. *Statist. Decisions*, 24(1):97–108, 2006.
- [32] Ludger Rüschendorf. *Mathematical risk analysis*. Springer Series in Operations Research and Financial Engineering. Springer, Heidelberg, 2013. Dependence, risk bounds, optimal allocations and portfolios.
- [33] Filippo Santambrogio. *Optimal transport for applied mathematicians*. Progress in Nonlinear Differential Equations and their Applications, 87. Birkhäuser/Springer, Cham, 2015. Calculus of variations, PDEs, and modeling.
- [34] C. Villani. *Topics in Optimal Transportation*, volume 58 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, 2003.
- [35] C. Villani. *Optimal Transport: Old and New*, volume 338 of *Grundlehren der mathematischen Wissenschaften*. Springer, 2009.

- [36] Bin Wang and Ruodu Wang. The complete mixability and convex minimization problems with monotone marginal densities. *J. Multivariate Anal.*, 102(10):1344–1360, 2011.